

AMENDMENTS TO THE CLAIMS

1. (Original) A method in a computing system for selecting web pages, comprising:

defining a set of web page attributes;

for each of a plurality of web pages,

automatically extracting from the web page one or more attribute values, each for one of the set of attributes;

storing the extracted attribute values together with a URL for the web page in a dimensional model of the plurality of web pages, by adding a row for the web page to a fact table of the dimensional model, the added fact table row containing the URL and referencing, for each attribute for which an attribute value was extracted, a row corresponding to the attribute value in a dimension table of the dimensional model corresponding to the attribute;

receiving a query specifying query attribute values for one or more of the set of attributes;

processing the query against the dimensional model, by:

for each of the set of attributes for which a query attribute value is specified, selecting the rows of the dimension table corresponding to the attribute that match the query attribute value;

joining the selected rows of the dimension tables corresponding to the attributes for which a query attribute value is specified to the fact table to produce a join result; and

generating a query result containing the URLs contained by the rows of the join result.

2. (Original) The method of claim 1 wherein the generated query result contains one or more attribute values extracted from the web pages corresponding to the URLs contained by the rows of the join result.

3. (Original) The method of claim 1 wherein the plurality of web pages comprises web pages that are company home pages, and wherein the defined set of attributes includes a company name attribute, a company type attribute, a company category attribute, and a company location attribute.

4. (Original) The method of claim 1, further comprising identifying the plurality of web pages by:

initializing the plurality of web pages by contributing one or more known web pages to the plurality of web pages; and

reiteratively, from one of the plurality of web pages:

following a link on the web page to a new web page; and

contributing the new web page to the plurality of web pages.

5. (Original) The method of claim 1, further comprising, for each of the plurality of web pages, for each attribute for which an attribute value was extracted:

determining whether an existing row of the dimension table corresponding to the attribute corresponds to the attribute value;

if an existing row of the dimension table corresponding to the attribute corresponds to the attribute value, including in the row added to the fact table a reference to the existing row of the dimension table;

if no existing row of the dimension table corresponding to the attribute corresponds to the attribute value:

adding a new row corresponding to the attribute value to the dimension table; and

including in the row added to the fact table a reference to the new row added to the dimension table.

6-26. (Canceled)

27. (Original) A method in a computing system for processing a search request against a dimensional model of a set of documents, the model comprising a fact table and two or more dimension table, the fact table being comprised of rows each containing a document reference and referencing, for each attribute for which an attribute value was extracted, a row corresponding to the attribute value in a dimension table of the dimensional model corresponding to the attribute, the method comprising:

receiving a search request specifying search request attribute tests for one or more of the set of attributes;

for each of the set of attributes for which a search request attribute test is specified, selecting the rows of the dimension table corresponding to the attribute that satisfy the search request attribute test;

joining the selected rows of the dimension tables corresponding to the attributes for which a search request attribute value is specified to the fact table to produce a join result; and

generating a search request result containing the document references contained by the rows of the join result.

28. (Original) The method of claim 27, wherein one of the specified search request attribute tests tests whether the attribute to which it corresponds matches a value specified by the search request.

29. (Original) The method of claim 27, wherein one of the specified search request attribute tests tests whether the attribute to which it corresponds is non-null.

30. (Original) The method of claim 27, wherein one of the specified search request attribute tests tests whether the attribute to which it corresponds falls within a range specified by the search request.

31. (Original) The method of claim 27, wherein one of the specified search request attribute tests tests whether the attribute to which it corresponds matches a pattern specified by the search request.

32. (Original) The method of claim 27, wherein one of the specified search request attribute tests tests whether the attribute to which it corresponds is among a list of alternative values specified by the search request.

33. (Original) The method of claim 27, wherein one of the specified search request attribute tests tests whether the attribute to which it corresponds satisfies a programmatic function specified by the search request.

34. (Original) The method of claim 27, wherein one of the specified search request attribute tests tests whether the attribute to which it corresponds satisfies a mathematical function specified by the search request.

35. (Original) The method of claim 27, further comprising:
performing a word search upon the generated search result; and
generating a second search result conveying the results of the performed word search.

36-49. (Cancelled)

50. (Original) The computer memories of claim 35 wherein each of the references to a document contained by the fact table rows is a reference to instructions for obtaining a document.

51-55. (Canceled)